Interdependent Privacy (IDP)

Mathias Humbert * and Kévin Huguenin

Synonyms

Collective privacy; group privacy; multi-party privacy; multi-subject privacy; networked privacy; peer privacy

Definitions

Interdependent privacy (IDP) refers to the fact that the actions or the data of one or multiple individuals have privacy implications for other individuals. Interdependent privacy risks usually stem from the fact that data features multiple individuals or that data that seemingly involves only certain individuals in fact reveals information about other individuals (typically because of correlation).

Background

The term "interdependent privacy" was coined by Biczók and Chia in 2013, probably inspired by that of "interdependent security" (see Laszka et al (2014)). Yet, interdependent privacy risks and associated solutions have been studied before that: See Humbert et al (2020) for a comprehensive survey on the topic.

The root cause of interdependent privacy is either that the disclosed data directly features multiple individuals or that the personal attributes of different individuals, hence the data associated with them, are correlated. Correlation is particularly strong when these individuals are related (e.g., friends, family members, coworkers). Interdependent privacy risks are often formalized through the notions of accessibility/visibility and inferrability of personal information.

* corresponding author

Correlation often indicates a predictive relationship between the personal attributes of individuals; these relationships can be exploited, by an adversary, at the expense of the individuals' privacy. Indeed, by considering the statistical properties of the relationship between the data of different individuals, the unknown personal attributes of an individual can be inferred, or, more generally, the knowledge about such attributes can be refined, algorithmically from the data disclosed by other individuals. Beyond correlation, interdependent privacy risks can also arise when individuals disclose, intentionally or not, to third parties the information related to another individual, which they were entrusted with. Interdependent privacy risks materialize for different types of personal data and information. For instance:

- The personal attributes of an individual can be inferred from those of their contacts (Jia et al 2017) as individuals tend to bond with individuals with whom they share certain characteristics (e.g., gender, age, preferences), a well-documented phenomenon known as homophily. Information about both these personal attributes and social relationships between individuals are available on online social networks. Moreover, users of online social networks can leak directly (i.e., without the need for inference) personal information about their contacts to third parties, including other individuals and app/service providers (Biczók and Chia 2013) (see User privacy research on online social networks).
- Multimedia data, such as photos and videos, create interdependent privacy risks as they often feature individuals other than the individual who captured them (Such and Criado 2018). Such content is commonly shared on online social networks and media (see User privacy research on online social networks).
- The location information of an individual can be inferred from that of their contacts, especially if co-location information is available (Olteanu et al 2017). Sources of co-location information include posts (with tags and mentions such as "with Alice") made on online social networks, photos, and IP addresses (see Location information (Privacy of) and, more specifically, Interdependent Location Privacy).
- The genetic material of an individual is inherited from their parents and (partially) passed on to their children. Consequently, the genomic information of an individual is correlated with that of their parents and children, but also, by extension, with those of their family members at large. Therefore, this information can be inferred from the information of those members. Direct-to-consumer genetic testing services enable individuals to obtain their genomic data (from a biological sample such as saliva) that they can subsequently upload on online platforms for various purposes, such as helping genomic research (see Genome privacy).

Standard privacy metrics are used to quantify privacy interdependent privacy risks (see Privacy metrics). Finally, data correlation can partially void the guarantees provided by differential privacy (see Differential Privacy and, more specifically, Pufferfish privacy) as it violates the core underlying assumption of data independence.

Theory

Studies on interdependent privacy scenarios rely on a number of standard theories and techniques. Note that these theories and techniques were not necessarily introduced in the context of interdependent privacy.

Statistical Inference

Statistical inference, and more specifically Bayesian inference, is used in interdependent privacy scenarios to update the knowledge about some personal attribute, typically of the target individual, given the observation of other attributes, typically those of other individuals. It exploits the statistical dependencies that exist between the target individual's data and that of their contacts (e.g., friends, family members, coworkers). Many works on interdependent privacy risk assessment rely on Bayesian inference (Jia et al 2017; Humbert et al 2017; Olteanu et al 2017).

Bayesian inference often relies on probabilistic graphical models (PGMs) for representing dependencies. Bayesian networks are used to represent directed dependencies, i.e, conditional probability distributions between random variables (see Fig. 1), whereas Markov networks are used for undirected dependencies, that is, joint probability distributions between random variables. Hidden Markov models (HMMs) are a special instance of Bayesian networks.

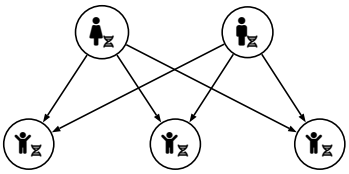

Fig. 1 Bayesian network used for genome inference in a family tree with two parents and three children. Source: Humbert et al (2017).

Game Theory

Game theory is used to study the interaction between multiple rational decision makers (referred to as *players*) who seek to optimize their utility in situations where the actions of a player can affect the utility of other players. A key concept of game theory is that of *equilibrium* (e.g., Nash equilibrium) that models stable states where, given other players' strategies, no player has an incentive to singlehandedly deviate from their strategy. As such, game theory is a first-class tool for predicting the strategy adopted by (rational) individuals in interdependent (privacy) scenarios. Many works on interdependent privacy (but also security) rely on game theory (Biczók and Chia 2013).

Cryptography and Access Control

Cryptography is used for building protection mechanisms in interdependent privacy scenarios, e.g., in case of sharing conflict on co-owned data, by requiring the consent of all co-owners before sharing this data or by hiding parts of the data. For this purpose, secret-sharing techniques are

particularly well suited. Attribute-based encryption and homomorphic encryption also represent helpful tools for designing cryptographic solutions in the context of interdependent privacy. We refer the reader to Secret Sharing Schemes, Attribute based encryption, and Homomorphic Encryption for more detail.

Access-control techniques are particularly useful in interdependent privacy scenarios as they help determine the audience of co-owned data (typically shared online) by granting or denying access to the data. For instance, relationship-based access control (ReBAC) can be used in the context of online social networks where social relationship information is available. Such schemes improve on Role-Based Access Control (RBAC) by adding granularity and context in the definition of roles. We refer the reader to Access Control Policies, Models, and Mechanisms.

Communication Privacy Management (CPM) Theory

Communication privacy management (CPM) theory models how individuals disclose personal information to others, building on the metaphor of boundaries that determine one's openness or closedness to the public. CPM theory is based on five key principles: ownership, control, rules, co-ownership, and boundary turbulence. Personal information belongs to a specific individual, referred to as the owner, who has right to control the information. When the owner shares personal information with another individual (a co-owner), collective boundaries are created. Privacy rules are then defined to coordinate the collective boundaries of co-owners. These rules control who can access the information and determine the possibility of extending the set of co-owners (linkage rules), the degree of access to the information (permeability rules), and the implication of co-owners in definitions of rules (ownership rules). The last principles, boundary turbulences, is related to scenarios when a co-owner discloses the information outside the initially defined boundaries, either intentionally or by mistake; this concept is particularly relevant for interdependent privacy scenarios. Several works on interdependent privacy rely on communication privacy management theory.

Open Problems and Future Directions

Interdependent privacy has been studied by different communities (including information security and privacy, data science, human-computer interaction/computer-supported cooperative work, and information systems), mostly in isolation. One of the reasons for this is that these different communities refer to the same concept using different terminologies, such as interdependent privacy, multi-party privacy, and networked privacy. One key challenge is to make these different research communities work closer with each other and collaborate to develop holistic solutions.

Most of the proposed inference algorithms and solutions are ad-hoc and data specific. In fact, many existing works focus on photos. Additionally, only a few works consider service providers as

a potential adversary: they focus only on protecting individuals' privacy with respect to other individuals. Overall, none of the existing solutions is generic enough to handle various data types and adversaries. Finally, interdependent privacy should also be tackled from a legal perspective, in the light of the various data protection laws across the world.

Cross-References

Access Control Policies, Models, and Mechanisms; Role-Based Access Control; Secret Sharing Schemes; Attribute based encryption; Location information (Privacy of); Genome privacy; Homomorphic Encryption; Privacy metrics; Differential Privacy; Pufferfish privacy; The Economics of Privacy; Interdependent Location Privacy; User privacy research on online social networks.

References

Biczók G, Chia PH (2013) Interdependent Privacy: Let Me Share Your Data. In: Proc. of FC, DOI 10.1007/978-3-642-39884-1_29

Humbert M, Ayday E, Hubaux JP, Telenti A (2017) Quantifying Interdependent Risks in Genomic Privacy. ACM Trans on Privacy and Security 20(1), DOI 10.1145/3035538

Humbert M, Trubert B, Huguenin K (2020) A survey on interdependent privacy. ACM Computing Surveys 52(6):122, DOI 10.1145/3360498

Jia J, Wang B, Zhang L, Gong NZ (2017) AttriInfer: Inferring User Attributes in Online Social Networks Using Markov Random Fields. In: Proc. of WWW, DOI 10.1145/3038912.3052695

Laszka A, Felegyhazi M, Buttyan L (2014) A Survey of Interdependent Information Security Games. ACM Computing Surveys 47(2), DOI 10.1145/2635673

Olteanu AM, Huguenin K, Shokri R, Humbert M, Hubaux JP (2017) Quantifying Interdependent Privacy Risks with Location Data. IEEE Trans on Mobile Computing 16(3), DOI 10.1109/TMC.2016.2561281

Such JM, Criado N (2018) Multiparty privacy in social media. Communications of the ACM 61(8), DOI 10.1145/3208039